ViVid: Depicting Dynamics in Stylized Live Photos

Amir Semmo
Hasso Plattner Institute for Digital Engineering, University of Potsdam

Max Reimann
Digital Masterpieces GmbH, Germany

Mandy Klingbeil
Digital Masterpieces GmbH, Germany

Sumit Shekhar
Hasso Plattner Institute for Digital Engineering, University of Potsdam

Matthias Trapp
Hasso Plattner Institute for Digital Engineering, University of Potsdam

Jürgen Döllner
Hasso Plattner Institute for Digital Engineering, University of Potsdam

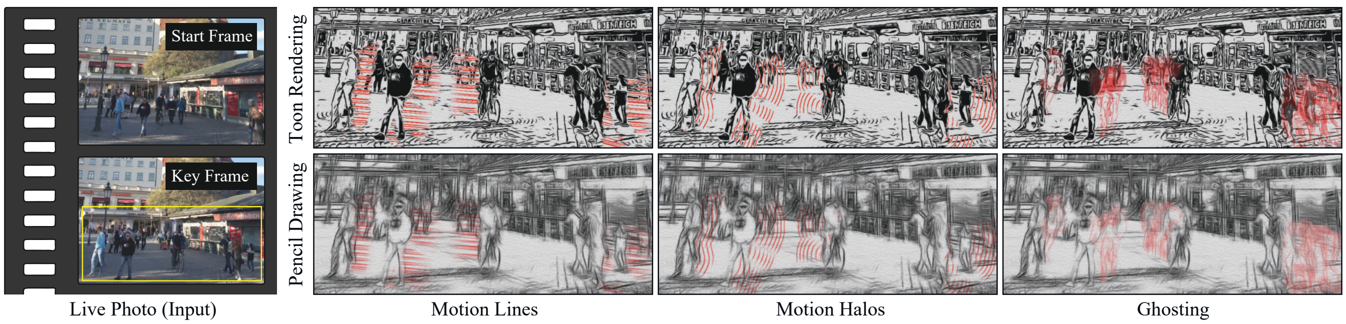

Figure 1: Overview of the rendering techniques implemented in *ViVid*. Our app enables to stylize Live Photos with a cartoon or pencil-drawing look that depict dynamics via motion lines, halos, or ghosting (highlighted in red for visualization purposes).

ABSTRACT

We present *ViVid*, a mobile app for iOS that empowers users to express dynamics in stylized Live Photos. This app uses state-of-the-art computer-vision techniques based on convolutional neural networks to estimate motion in the video footage that is captured together with a photo. Based on this analysis and best practices of contemporary art, photos can be stylized as a pencil drawing or cartoon look that includes design elements to visually suggest motion, such as ghosts, motion lines and halos. Its interactive parameterizations enable users to filter and art-direct composition variables, such as color, size and opacity. *ViVid* is based on Apple's CoreML, Metal and PhotoKit APIs for optimized on-device processing. Thus, the motion estimation is scheduled to utilize the dedicated neural engine, while shading-based image stylization is able to process the video footage in real-time on the GPU. This way, the app provides a unique tool for creating lively photo stylizations with ease.

CCS CONCEPTS

• **Computing methodologies** → **Non-photorealistic rendering**; **Image processing**; • **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**;

KEYWORDS

depicting dynamics, mobile devices, motion, stylization, live photos

SIGGRAPH '19 Appy Hour, July 28 - August 01, 2019, Los Angeles, CA, USA

ACM Reference Format:
Amir Semmo, Max Reimann, Mandy Klingbeil, Sumit Shekhar, Matthias Trapp, and Jürgen Döllner. 2019. ViVid: Depicting Dynamics in Stylized Live Photos. In *Proceedings of SIGGRAPH '19 Appy Hour*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3305365.3329726

1 MOTIVATION

Image filters, particularly those used for mobile expressive rendering, have become pervasive tools in casual creativity applications [Dev 2013] and for users that seek to increase the viewers' engagement [Bakhshi et al. 2015]. These filters, however, typically only operate in the spatial domain, thus excluding fundamental temporal-related aspects of pictorial semiotics [Rudner 1951] for expressing motion [Nienhaus and Döllner 2005]. With the continuous advancements in mobile camera hardware, capturing multi-dimensional data has become a common feature—e.g., Apple's Live Photos captures the 1.5 second video footage before and after a photo—, but so far this information has not been actively utilized in mobile expressive rendering apps, e.g., to depict dynamics.

In this work, we present *ViVid*, a mobile app that enables users to utilize both spatial and temporal information to express dynamics in a stylized Live Photo. At this, state-of-the-art convolutional neural networks are used to estimate movement in the video footage and art-direct visual motion cues in the stylized photo. For this purpose, we employ best practices of contemporary art [Cutting 2002; Walker 2015] and illustrations (e.g., comics [McCloud 2006, 2008]) to visually depict dynamics, such as using motion lines, halos, and ghosts as first-class design entities (Figure 1).

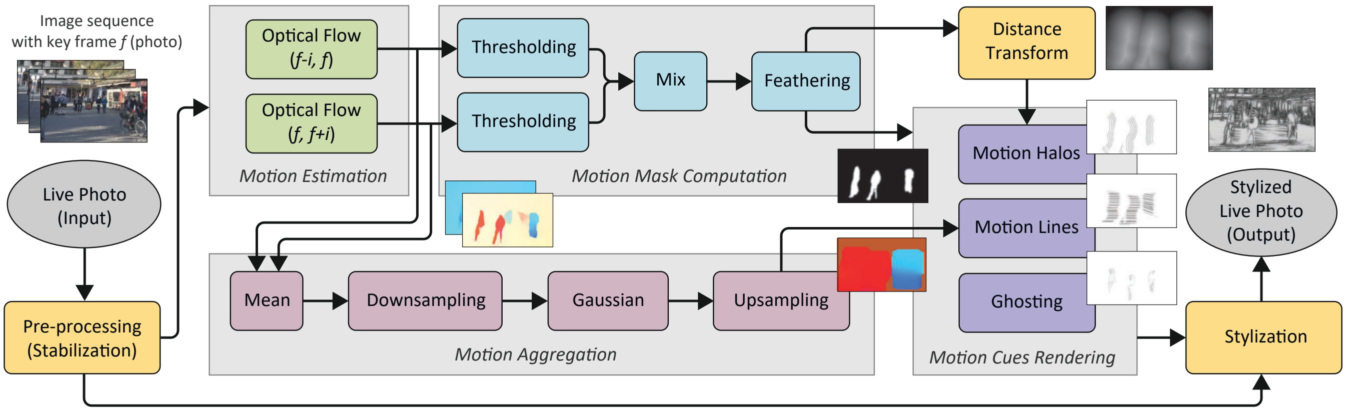

Figure 2: Schematic overview of the processing pipeline implemented in our app *ViVid*.

2 TECHNICAL APPROACH

The main challenge for depicting dynamics is to detect and separate moving from non-moving objects. This is a challenging task, because camera shake is typically introduced when taking a photo, and optical image stabilization cannot ultimately compensate it. Previous works approached this problem by using background subtraction that thresholds Euclidean distances of pixel colors [Nienhaus et al. 2008] or feature tracking with homography projection [Collomosse et al. 2003] to compensate camera motion. In this work, we also align images features via homography projections but then use a CNN-based model for motion estimation that does not depend on explicit feature tracking. An overview of our processing pipeline is shown in Figure 2 and comprises the following stages:

Pre-processing (Stabilization). We employ Apple's Vision framework to estimate the homography between video frames and use the perspective warp matrix to align images features for stabilization.

Motion Estimation. We ported the PWC-Net [Sun et al. 2018] to CoreML to utilize the dedicated neural engine for optical flow computation. In particular, the CNN-based architecture is able to estimate large displacement flow as demonstrated with the KITTI 2015 benchmarks.

Motion Mask Computation. The optical flow is computed for video frames before and after a selected key frame to detect moving objects in transition. The information can also be combined with a user-defined object mask for refinement.

Motion Aggregation. Motion cues may need to be aligned with the dominant direction of moving objects, thus we follow a simple strategy for motion aggregation by low-pass filtering the downsampled mean of the optical flow and using the upsampled result.

Motion Cues Rendering. We implemented three shading techniques for rendering motion cues: motion halos that threshold the Euclidean distance map of the computed motion mask, motion lines that align 2D textures with the motion path, and ghosts depicting silhouettes of moving objects.

Stylization. We implemented the techniques by Winnemöller et al. [Winnemöller et al. 2012, 2006] for toon rendering and the technique by Lu et al. [Lu et al. 2012] to obtain sketchy drawings.

We are able to obtain motion estimations in 520ms for 1024×448 pixel images and 300ms for 640×384 pixel images on an iPhone XS. The shading-based motion cue rendering and stylization stages run in real-time by employing Apple's CoreML and Metal APIs. This way, interactive parameterization of these stages enable users to filter and art-direct composition variables, such as the color, size, and opacity of motion cues.

ACKNOWLEDGEMENTS

This work was funded by the Federal Ministry of Education and Research (BMBF), Germany, for the AVA project 01IS15041.

REFERENCES

Saeideh Bakhshi, David A Shamma, Lyndon Kennedy, and Eric Gilbert. 2015. Why We Filter Our Photos and How It Impacts Engagement. In *Proc. ICWSM*. AAAI Press, 12–21.

John P. Collomosse, David Rowntree, and Peter M. Hall. 2003. Cartoon-Style Rendering of Motion from Video. In *Proc. Vision, Video and Graphics (VVG)*. The Eurographics Association, 117–124. DOI: http://dx.doi.org/10.2312/vvg.20031016

James E Cutting. 2002. Representing Motion in a Static Image: Constraints and Parallels in Art, Science, and Popular Culture. *Perception* 31, 10 (2002), 1165–1193. DOI: http://dx.doi.org/10.1068/p3318

Kapil Dev. 2013. Mobile Expressive Renderings: The State of the Art. *IEEE Computer Graphics and Applications* 33, 3 (May/June 2013), 22–31. DOI: http://dx.doi.org/10.1109/MCG.2013.20

Cewu Lu, Li Xu, and Jiaya Jia. 2012. Combining Sketch and Tone for Pencil Drawing Production. In *Proc. NPAR*. The Eurographics Association, 65–73. DOI: http://dx.doi.org/10.2312/PE/NPAR/NPAR12/065-073

Scott McCloud. 2006. *Making Comics: Storytelling Secrets of Comics, Manga and Graphic Novels*. Harper New York.

Scott McCloud. 2008. *Understanding Comics: The Invisible Art*. Paw Prints.

Marc Nienhaus and Jürgen Döllner. 2005. Depicting Dynamics Using Principles of Visual Art and Narrations. *IEEE Computer Graphics and Applications* 25, 3 (2005), 40–51. DOI: http://dx.doi.org/10.1109/MCG.2005.53

Marc Nienhaus, Holger Winnemöller, Jürgen Döllner, and Bruce Gooch. 2008. Forward Lean - Deriving Motion Illustrations From Video. In *Proc. SIGGRAPH Asia Sketches*.

Richard Rudner. 1951. On Semiotic Aesthetics. *The Journal of Aesthetics and Art Criticism* 10, 1 (Sept. 1951), 67–77. DOI: http://dx.doi.org/10.2307/426789

D. Sun, X. Yang, M. Liu, and J Kautz. 2018. PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume. In *Proc. IEEE CVPR*. IEEE, 8934–8943. DOI: http://dx.doi.org/10.1109/CVPR.2018.00931

Peter Walker. 2015. Depicting Visual Motion in Still Images: Forward Leaning and a Left to Right Bias for Lateral Movement. *Perception* 44, 2 (2015), 111–128. DOI: http://dx.doi.org/10.1068/p7897

Holger Winnemöller, Jan Eric Kyprianidis, and Sven Olsen. 2012. XDoG: An eXtended Difference-of-Gaussians Compendium including Advanced Image Stylization. *Computers & Graphics* 36, 6 (Oct. 2012), 740–753. DOI: http://dx.doi.org/10.1016/j.cag.2012.03.004

Holger Winnemöller, Sven C. Olsen, and Bruce Gooch. 2006. Real-Time Video Abstraction. *ACM Transactions on Graphics* 25, 3 (July 2006), 1221–1226. DOI: http://dx.doi.org/10.1145/1141911.1142018